Claims

1. A zinc finger polypeptide which binds to a target DNA sequence containing a modified base but not to an identical sequence containing the equivalent unmodified base.

2. A polypeptide according to claim 1, wherein the target DNA sequence comprises a triplet having 5-meC at the central position, and binding to the 5-meC residue by an α-helical zinc finger binding motif in the polypeptide is achieved by placing an Ala residue at position +3 of the α-helix.

3. A method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a DNA triplet in target DNA sequence comprising 5-meC as the central residue in the target DNA triplet, wherein binding to the 5-meC residue by an α-helical zinc finger DNA binding motif of the polypeptide is achieved by placing an Ala residue at position +3 of the α-helix of the zinc finger.

4. A method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a DNA triplet in target DNA sequence comprising 5-meC, but not to an identical triplet comprising unmethylated C, wherein binding to each base of the triplet by an α-helical zinc finger DNA binding motif in the polypeptide is determined as follows:

a) if the 5' base in the triplet is G, then position +6 in the α-helix is Arg and/or position ++2 is Asp;

b) if the 5' base in the triplet is A, then position +6 in the α-helix is Gln or Glu and ++2 is not Asp;

c) if the 5' base in the triplet is T, then position +6 in the α-helix is Ser or Thr and position ++2 is Asp; or position +6 is a hydrophobic amino acid other than Ala;

d) if the 5' base in the triplet is C, then position +6 in the α-helix may be any amino acid, provided that position ++2 in the α-helix is not Asp;

e) if the central base in the triplet is G, then position +3 in the α-helix is His;

f) if the central base in the triplet is A, then position +3 in the α-helix is Asn;

g) if the central base in the triplet is T, then position +3 in the α-helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;

h) if the central base in the triplet is 5-meC, then position +3 in the α-helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;

i) if the 3' base in the triplet is G, then position -1 in the α-helix is Arg;

j) if the 3' base in the triplet is A, then position -1 in the α-helix is Gln and position +2 is Ala;

k) if the 3' base in the triplet is T, then position -1 in the α-helix is Asn; or position -1 is Gln and position +2 is Ser;

l) if the 3' base in the triplet is C, then position -1 in the α-helix is Asp and Position +1 is Arg.

5. A method for producing a zinc finger polypeptide capable of binding to a DNA sequence comprising a modified residue, but not to an identical sequence comprising an equivalent unmodified residue, comprising the steps of:

a) providing a DNA library encoding a repertoire of zinc finger polypeptides, the DNA members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of an α-helical zinc finger binding motif of the zinc finger polypeptides;

b) displaying the library in a selection system and screening it against a target DNA sequence comprising the modified residue;

c) isolating the DNA members of the library encoding zinc finger polypeptides capable of binding to the target sequence; and

d) optionally, verifying that the zinc finger polypeptides do not bind significantly to a DNA sequence identical to the target DNA sequence but containing the equivalent unmodified residue in place of the modified residue.

6. A method according to claim 5, wherein the nucleic acid library encodes a repertoire of zinc finger polypeptides each possessing more than one zinc fingers, the nucleic acid members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of the α-helix in a zinc finger and at one or more of the positions encoding residues -1, 2, 3 and 6 of the α-helix in a further zinc finger of the zinc finger polypeptides.

7. A method according to claim 5 or claim 6, wherein the modified residue is 5-meC and the unmodified residue is C.

8. A method according to claim 5 or claim 6, wherein the modified residue is U and the unmodified residue is T.

9. A method according to any one of claims 5 to 8, wherein the library is screened by phage display.

10. A method according to any one of claims 5 to 9, wherein the or each zinc finger has the general primary structure

(A) X^a **C** X_{2-4} **C** X_{2-3} F X^c X X X X L X X **H** X X X^b **H** - linker

-1 1 2 3 4 5 6 7 8 9

wherein X (including X^a, X^b and X^c) is any amino acid.

11. A method according to claim 10 wherein X^a is $^{F}/_{Y}$-X or P-$^{F}/_{Y}$-X.

12. A method according to claim 10 or claim 11 wherein X_{2-4} is selected from any one of: S-X, E-X, K-X, T-X, P-X and R-X.

13. A method according to any one of claims 10 to 12 wherein X^b is T or I.

14. A method according to any one of claims 10 to 13 wherein X_{2-3} is G-K-A, G-K-C, G-K-S, G-K-G, M-R-N or M-R.

15. A method according to any one of claims 10 to 14 wherein the linker is T-G-E-K or T-G-E-K-P.

16. A method according to any one of claims 10 to 15 wherein position +9 is R or K.

17. A method according to any one of claims 10 to 16 wherein positions +1, +5 and +8 are not occupied by any one of the hydrophobic amino acids, F, W or Y.

18. A method according to claim 17 wherein positions +1, +5 and +8 are occupied by the residues K, T and Q respectively.

19. A method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a DNA triplet in target DNA sequence comprising 5-meC, but not to an identical triplet comprising unmethylated C:

a) selecting a model zinc finger domain from the group consisting of naturally occurring zinc fingers and consensus zinc fingers; and

b) mutating the finger by the method of any one of claims 3 to 17.

20. A method according to claim 19, wherein the model zinc finger is a consensus zinc finger whose structure is selected from the group consisting of the consensus structure P Y K C P E C G K S F S Q K S D L V K H Q R T H T G, and the consensus structure P Y K C S E C G K A F S Q K S N L T R H Q R I H T G E K P.

21. A method according to claim 19 wherein the model zinc finger is a naturally occurring zinc finger whose structure is selected from one finger of a protein selected from the group consisting of Zif 268 (Elrod-Erickson *et al.*, (1996) Structure 4:1171-1180), GLI (Pavletich and Pabo, (1993) Science 261:1701-1707), Tramtrack (Fairall *et*

al., (1993) Nature 366:483-487) and YY1 (Houbaviy *et al.*, (1996) PNAS (USA) 93:13577-13582).

22. A method according to claim 21 wherein the model zinc finger is finger 2 of Zif 268.

23. A method according to any one of claims 3 to 22 wherein the binding protein comprises two or more zinc finger binding motifs, placed N-terminus to C-terminus.

24. A method according to claim 22, wherein the N-terminal zinc finger is preceded by a leader peptide having the sequence MAEEKP.

25. A method according to claim 23 or claim 24, wherein the DNA binding protein is constructed by recombinant DNA technology, the method comprising the steps of:

a) preparing a DNA coding sequence encoding two or more zinc finger binding preparable according to claim 23 or 24, placed N-terminus to C-terminus;
b) inserting the DNA sequence into a suitable expression vector; and
c) expressing the DNA sequence in a host organism in order to obtain the DNA binding protein.

26. A method according to one of claims 3 to 25 comprising the additional steps of subjecting the DNA binding protein to one or more rounds of randomisation and selection in order to improve the characteristics thereof.

27. A zinc finger polypeptide which binds to a target DNA sequence containing a modified base but not to an identical sequence containing the equivalent unmodified base, preparable by a method according to any one of claims 3 to 26.